

Citation for published version:

Nikolopoulos, K, Syntetos, A, Boylan, JE, Petropoulos, F & Assimakopoulos, V 2011, 'An Aggregate - Disaggregate Intermittent Demand Approach (ADIDA) to forecasting: an empirical proposition and analysis', *Journal of the Operational Research Society*, vol. 62, no. 3, pp. 544-554. <https://doi.org/10.1057/jors.2010.32>

DOI:

[10.1057/jors.2010.32](https://doi.org/10.1057/jors.2010.32)

Publication date:

2011

Document Version

Early version, also known as pre-print

[Link to publication](#)

University of Bath

Alternative formats

If you require this document in an alternative format, please contact:
openaccess@bath.ac.uk

General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

An Aggregate - Disaggregate Intermittent Demand Approach (ADIDA) to Forecasting: An Empirical Proposition and Analysis

(Submitted for peer review for the *Journal of the Operational Research Society*)

K Nikolopoulos¹, AA Syntetos^{2*}, JE Boylan³, F Petropoulos⁴, V Assimakopoulos⁵

¹University of Manchester, UK; ²University of Salford, UK; ³Buckinghamshire New University, UK;

⁴National Technical University of Athens, Greece; ⁵Ministry of Economy and Finance, Greece

Abstract

Intermittent demand patterns are characterised by infrequent demand arrivals coupled with variable demand sizes. Such patterns prevail in many industrial applications, including IT, automotive, aerospace and military. An intuitively appealing strategy to deal with such patterns from a forecasting perspective is to aggregate demand in lower-frequency ‘time buckets’ thereby reducing the presence of zero observations. However, such aggregation may result in losing useful information, as the frequency of observations is reduced. In this paper, we explore the effects of aggregation by investigating 5,000 Stock Keeping Units (SKUs) from the Royal Air Force (RAF, UK). We are also concerned with the empirical determination of an optimum aggregation level as well as the effects of aggregating demand in time buckets that equal the lead time length (plus review period). This part of the analysis is of direct relevance to a (periodic) inventory management setting where such cumulative lead-time demand estimates are required. Our study allows insights to be gained into the value of aggregation in an intermittent demand context. The paper concludes with an agenda for further research in this area.

Keywords: Demand Forecasting; Inventory Management; Intermittent Demand; Aggregation; Empirical Investigation

Acknowledgements:

The work conducted by Aris Syntetos and John Boylan has been funded by the Engineering and Physical Sciences Research Council (EPSRC, UK) grant no. EP/F012632/1 (a project entitled: *Forecasting and inventory management: bridging the gap*). More information on this project may be obtained at: http://www.mams.salford.ac.uk/CORAS/Projects/Bridging_the_Gap/.

* Corresponding author: Aris A. Syntetos, e-mail: a.syntetos@salford.ac.uk, Tel. +44 (0)161 295 5804
Centre for Operational Research and Applied Statistics, Salford Business School, University of Salford,
Maxwell Building, Manchester M5 4WT, UK

1. Introduction

Intermittent demand for products appears sporadically, with some time periods showing no demand at all. When demand occurs, the demand size may be variable, perhaps highly so. Intermittent demand items may be any Stock Keeping Unit (SKU) within the range of products offered by an organization at any level of the supply chain. Such items may collectively account for up to 60% of the total stock value (Johnston *et al*, 2003) and are particularly prevalent in the aerospace, automotive, military and IT sectors. They are often the items at greatest risk of obsolescence. Inventory control decisions for intermittent items are needed to determine inventory replenishment rules. These decisions can be made more intelligently if supported by more accurate demand forecasts. Improvements in forecasting and stock control may be translated to significant reductions in wastage or scrap, and very substantial cost savings.

Replenishment requirements should be calculated according to the anticipated probability distribution of demand over the lead-time. However, single-period forecasts are often aggregated over lead-time using *ad hoc* formulae in forecasting software, and there is limited guidance on the issue in the academic literature. Hence, there is a need to address the question of *forecast horizon aggregation*, to design coherent aggregation mechanisms. *Temporal aggregation* refers to aggregation in which a low frequency time series (e.g. quarterly) is derived from a high frequency time series (e.g. monthly) and is used for forecasting purposes. This is ignored in much commercial practice, and there is only a small body of academic research on the subject. Nevertheless, temporal aggregation is a promising approach for intermittent demand, as forecasts at higher levels of aggregation

are generally more accurate and less variable than those at lower levels of aggregation. The level of temporal aggregation may be chosen to mirror that of the forecast horizon (lead-time), or may exceed it, in which case disaggregation mechanisms are required.

An obvious disadvantage related to temporal aggregation is that of losing information since the frequency and number of observations is reduced. However, the accumulation of demand observations in lower-frequency ‘time buckets’ reduces the number of zero demands and the resulting series bear a greater resemblance to those for fast-moving items. As discussed above, requirements for such SKUs are easier to estimate and they are typically associated with lower errors than those resulting from forecasting ‘slow’/intermittent demands. The effect of the level of aggregation on forecast accuracy will be investigated further and empirically optimum levels will be identified.

An inherent difficulty associated with intermittent demand patterns relates to the identification of the underlying series’ characteristics such as trend and seasonality. The presence of zeroes precludes the accurate estimation of such components (that are hidden at the high frequency level) and *aggregation* should facilitate, theoretically at least, such a process. The issue of trend and seasonal detection is not discussed further here, but is part of a larger research programme being undertaken by the authors of this paper.

1.1 Research objectives

Exploring the effects of aggregation empirically constitutes the main objective of our paper. Determining the (potential) optimum aggregation levels constitutes a further objective. This issue is also analysed empirically and appropriate cut-off points are

specified. Our analysis is based on 5,000 SKUs from the Royal Air Force (RAF, UK). The issue of disaggregating forecasts is considered as well and linked to the theoretical properties of the original series.

Intermittent demand estimates are typically associated with a high variance due to the two sources of variation (demand arrivals and demand sizes). It is therefore of interest to investigate the effect of the aggregation approach on the sampling error of the mean (as reflected on the Mean Squared Error – MSE), and on the bias of mean estimates. In addition, it is certainly worthwhile exploring the performance of estimators that were originally designed for fast-moving products and contrasting their accuracy to that obtained from intermittent demand Croston-based estimators.

Finally, the effects of aggregating demand in buckets that equal the lead time length (plus review period) are investigated. This analysis is of direct relevance to a (periodic) inventory management setting where cumulative demand estimates over that time horizon are required.

In summary, the objectives of our empirical study are as follows:

1. To provide for the first time some results on the performance of Temporal Aggregation when used for items with intermittent demands;
2. To empirically determine optimum aggregation levels;
3. To consider appropriate disaggregation mechanisms and link their performance to the statistical properties of the original series;

4. To assess the effects of temporal aggregation in time buckets that equal the lead time (plus review period).

As discussed above, our analysis is conducted on a dataset from the RAF that consists of the individual demand histories of 5,000 SKUs. Demand has been recorded monthly and the data available covers 7 years' history (84 monthly demand observations). The actual lead-time is available for each of those SKUs.

1.2 Structure of the paper

The remainder of our paper is organised in five further sections. In the next Section a literature review is presented, followed, in Section 3, by the conceptual development of our approach, a demonstration of how the approach may be applied in practice and a discussion on operational issues. Section 4 covers some basic information regarding the dataset used for the purposes of our investigation and details related to the structure of our experiment. The empirical results are analysed in Section 5 and, finally, the conclusions of our work along with some natural next steps of research are presented in Section 6.

2. Research background

The methods that are employed in practice to forecast intermittent demand requirements are often quite straightforward. Single Exponential Smoothing (SES) has long been known to produce biased forecasts immediately after a demand occurrence (Croston, 1972). This is important operationally, as stock requirements are often recalculated at this point. Croston (1972) proposed a method that captures the compound nature of the underlying

demand structure. In particular he suggested using SES for separately forecasting the interval between demand incidences, and the demand sizes. The ratio of the latter to the former may then be used to estimate the mean demand per time period. However, Croston's method has also been shown to be biased (Syntetos and Boylan, 2001), and an approximately unbiased variation has since been developed (Syntetos and Boylan, 2005). Known as the Syntetos-Boylan Approximation (SBA), this estimator is calculated by multiplying Croston's estimate by $(1 - \alpha/2)$, where α is the smoothing constant used to update estimates of the mean interval between demands. This method and other variations yield similar accuracy benefits over Croston's method (Teunter and Sani, 2009). There have been no studies in the academic literature comparing the accuracy of these methods at different levels of temporal aggregation. With respect to stock control, estimation of demand variance must also be addressed. Temporal aggregation in blocks that equal the lead-time (LT) plus review period (R) was examined by Syntetos and Boylan (2006) who proposed a cumulative Mean Squared Error smoothing procedure for demand variance estimation. This is an important issue, but is not further pursued in our current study. In this research paper, we are solely concerned with mean level estimation.

A weakness of model-based forecasting methods is that a standard distribution, such as the Poisson or Negative Binomial, needs to be hypothesised. Real data often exhibit greater variation than standard distributions. To address this issue, Willemain *et al* (2004) introduced a bootstrapping method for intermittent demand. Their method is not model-based but instead is a heuristic that combines a Markov process, bootstrapping and 'jittering' to simulate an entire distribution for lead-time demand rather than a single

forecast. (Jittering is an ad-hoc procedure designed to allow greater variation than that already observed. The process enables the sampling of demand size values that have not been observed in the demand history.) An alternative non-parametric approach, based on Kernel Density Estimation (KDE) has been proposed by Boylan *et al* (2008). This also allows for the generation of demands not previously observed, but with greater flexibility regarding the ‘smoothing’ of the empirical distribution. Porras and Dekker (2008) proposed an *Empirical Method*, based on the construction of a histogram of demands over *LT*. This was found to yield lower inventory costs than Willemain’s bootstrapping method. However, it cannot extrapolate beyond previous demands, making it difficult to attain high service level targets.

The issue of *forecast horizon aggregation* has been addressed for non-intermittent demand forecasting, using models based on Multiple Source of Error (Johnston and Harrison, 1986) and Single Source of Error (Snyder *et al*, 1999). This issue is also addressed directly by the ‘Empirical Method’ discussed above.

3. An Aggregate-Disaggregate Intermittent Demand Approach (ADIDA) to forecasting

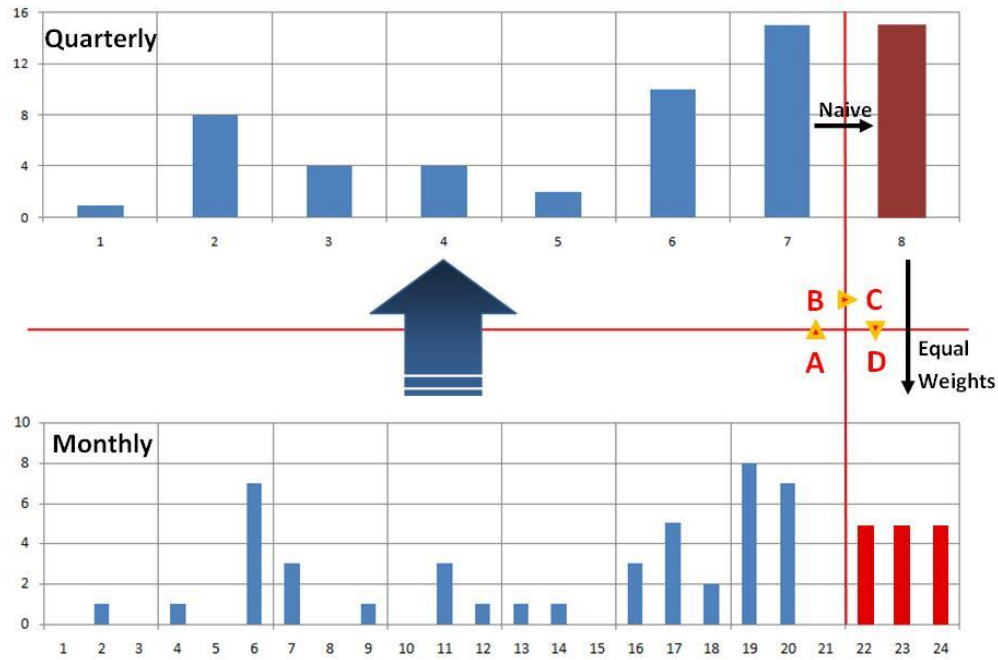
With respect to *temporal aggregation*, we must distinguish between overlapping and non-overlapping cases. In non-overlapping aggregation, the demand series are divided into consecutive non-overlapping blocks of equal length. In overlapping aggregation, the blocks are of equal lengths but, at each period, the oldest observation is dropped and the newest is included.

To address aggregation from a theoretical perspective, a modelling framework is required that is rigorous, flexible, and robust when applied to real data. Croston (1972) assumed that the time between demands was an i.i.d. (identically and independently distributed) process. There is empirical evidence in support of this assumption, including that provided in this paper (see Section 4). However, some series show evidence of auto-correlation (Willemain *et al*, 1994). Snyder (2002) pointed out that Croston's demand model is incompatible with his forecasting method, because exponential smoothing methods are not designed for i.i.d. demand. Nevertheless, one needs to consider the following: a method that is optimal for one particular model may be severely sub-optimal for another model. Syntetos *et al* (2006) argued that, for intermittent demand, robustness of a method across a wide range of possible underlying demand models is more important than optimality under one particular model. The scarcity of demand observations presents a significant inherent difficulty in identifying the demand model.

An alternative approach in modelling is to adopt Auto-Regressive Integrated Moving Average (ARIMA) demand models. Under such a modelling approach, temporal aggregation for non-intermittent time-series has been the subject of many research papers over the last four decades. Results for 'flow' (overlapping aggregation) and 'stock' (non-overlapping aggregation) have been obtained for ARIMA and ARIMAX models (ARIMA with exogenous variables), ARMA-GARCH (Generalized Auto-Regressive Conditional Heteroscedastic) and Vector (V)ARMA models. (See Silvestrini and Veredas (2008) for a comprehensive review of developments.) Results have also been obtained for Auto-Regressive Fractionally IMA (ARFIMA) models by Tsai and Chan (2005).

The literature on temporal aggregation of intermittent time-series is much less extensive. The most comprehensive analysis was conducted by Brännäs *et al* (2002). The authors investigated temporal aggregation of Integer Auto-Regressive (INAR) processes. Whilst overlapping aggregation of processes such as INAR (1) preserves the model form, the case of non-overlapping temporal aggregation is more complex. For example, non-overlapping aggregation of an INAR (1) model produces a model that resembles an INARMA (1,1) but with moving average components that are correlated.

We return to the issue of intermittent demand theoretical modelling for aggregation purposes in the last section of this paper where the next steps of our research are also discussed. In our work we are concerned with non-overlapping aggregation (for non-auto-correlated data). The process governing such an aggregation mechanism and the way it may be utilized for extrapolation purposes is pictorially presented in Figure 1.



KEY: **A:** Original data (months); **B:** Aggregate data (quarters); **C:** A quarterly forecast is produced;
D: The quarterly forecast is broken down to three equal monthly forecasts

Figure 1. ADIDA forecasting framework

Suppose the original data is reported in monthly time buckets and there are currently 21 historical observations available (part A - Figure 1). The monthly time series exhibit intermittence, which may be reduced by aggregating individual observations into e.g. quarterly data. The resulting series consists of 7 aggregated observations (part B - Figure 1) and constitutes what one may term ‘regular demand’. In this specific example, the resulting series is still quite volatile but we would expect a considerable variance reduction as compared to the original series.

Since the new series is non-intermittent we may use any method originally designed and/or practically utilised for fast demand items in order to extrapolate and obtain the one-step ahead quarterly forecast. The simplest choice would be to use the Naïve method although an empirical competition could be conducted among various estimators by holding out the 7th observation and fitting methods in the first six quarters. A whole range of methods becomes available under the present structure of the series whereas in the original series the choice would be limited among few alternatives (e.g. Croston’s method and the Syntetos-Boylan Approximation).

For the sake of illustration, we extrapolate the series with the Naïve method (part C - Figure 1) – i.e. the actual aggregate demand in the 7th quarter becomes the forecast for the unknown *cumulative* demand for the full 8th quarter. We intentionally use the term ‘cumulative’ as this aggregate forecast could be seen as a direct way to get a cumulative estimate for the next three months if this is what we require in operational terms (i.e. for stock control purposes). This issue is further discussed in Section 5 of the paper.

Returning to the original task of providing monthly forecasts, we need to break down (i.e. disaggregate) the quarterly forecast into three monthly forecasts for months 22, 23 and 24. The simplest option would be to disaggregate the forecast into monthly estimates using equal weights (i.e. in this case $1/3$ of the quarterly forecast). This phase is illustrated in part D of Figure 1, and this concludes the ADIDA process.

ADIDA is an intuitively appealing aggregation/disaggregation mechanism that aims at reducing (or eliminating) intermittence for the purpose of providing us with more forecasting tools. ADIDA is associated neither with a specific extrapolation method, nor with a specific disaggregation algorithm. Furthermore, the level of aggregation is an open question. Thus, ADIDA constitutes a generic framework offering an alternative way to produce forecasts. However, one important issue that needs to be mentioned is that the computational cost associated with this four-phase data-manipulating process compares unfavourably to that related to the application of a single extrapolative method.

In the next two sections we provide some empirical evidence on the effectiveness of ADIDA for intermittent demand time series forecasting.

4. Empirical data and experimental structure

The empirical database available for the purposes of our research consists of the individual monthly demand histories of 5,000 SKUs over 7 years (84 monthly demand observations, from Jan. 1996 to Dec. 2002 inclusive) from the Royal Air Force (RAF). The lead-time length and unit cost information are also provided for each of the SKUs. The same database has been used in an earlier study by Syntetos *et al* (2009a). Detailed descriptive

statistics (to the first decimal place) on the demand data series characteristics and lead-time (LT) + review (R) information ($R = 1$) are presented in Table 1.

| 5000 SKUs | Demand Sizes | | Demand Intervals | | Demand per period | | $LT+R$ |
|---------------|--------------|-------|------------------|-------|-------------------|-------|--------|
| | Mean | StDev | Mean | StDev | Mean | StDev | |
| Min. | 1.0 | 0.0 | 3.8 | 0.0 | 0.1 | 0.2 | 1.0 |
| 25%ile | 1.6 | 0.8 | 7.3 | 5.4 | 0.2 | 0.5 | 6.0 |
| Median | 3.8 | 3.1 | 9.0 | 6.9 | 0.4 | 1.5 | 10.0 |
| 75%ile | 11.3 | 9.3 | 11.6 | 8.6 | 1.2 | 4.4 | 13.0 |
| Max. | 668.0 | 874.4 | 24.0 | 16.5 | 65.1 | 275.7 | 34.0 |

Table 1. Demand data descriptive statistics

The lag-1 and lag-12 autocorrelation of the series is indicated in Figures 2a and 2b respectively. The series exhibit little autocorrelation and this will inform some interpretations of the empirical results discussed in sub-section 5.3.

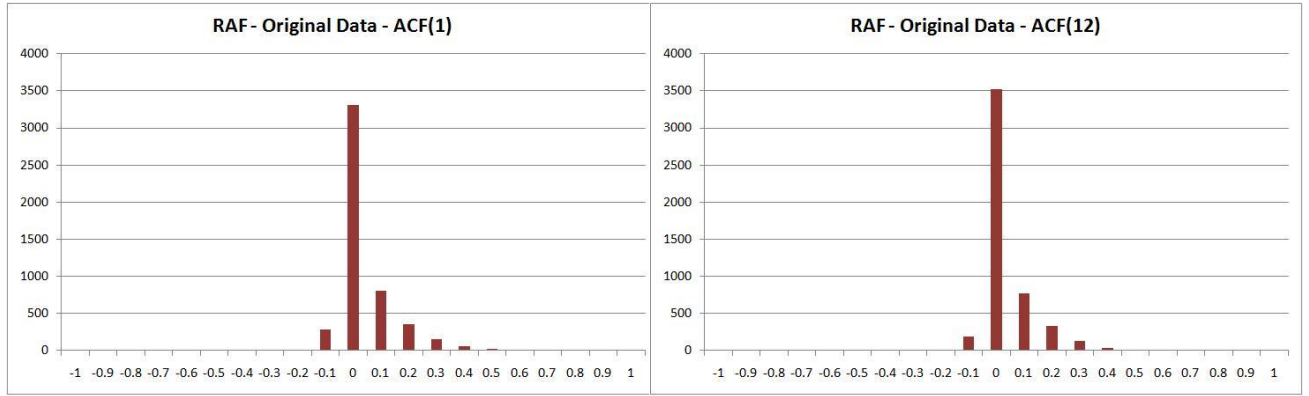


Figure 2. Auto-Correlation Function (ACF) – lag 1 and lag 12 (5,000 SKUs)

5. Empirical investigation

In this section we provide results from an empirical investigation that aims at:

- exploring (empirically) the optimum aggregation levels;
- assessing the performance of various disaggregation mechanisms;
- testing the effectiveness of ADIDA in terms of various forecast accuracy metrics.

5.1. Optimal aggregation levels (per series): an empirical determination

The first objective of our analysis is to identify, empirically, the optimal aggregation levels per series. In order to do so, we considered two forecasting methods: the Naïve and Syntetos-Boylan Approximation (SBA) and we used as the holdout period the last two years of each series (24 observations). We performed a sliding simulation (rolling evaluation) over this out-of-sample via producing one-step-ahead forecasts; thus we calculated 24 one-step-ahead errors for each series ($Error = Actual - Forecast$), for each of the forecasting methods considered. Absolute scaled Errors were subsequently calculated (AsE, Hyndman and Koehler, 2006); the mean value of this metric - MAsE, across each series, is reported in Figure 5 for six randomly selected series.

In order to get forecasts via the ADIDA framework we started creating buckets of aggregated data.

Aggregation level = 1: The data remain unchanged. Thus forecasts are produced via normal extrapolation with either Naïve or SBA on the original data. The implementation of SBA relies upon Single Exponential Smoothing (SES) forecasts of the demand sizes (when demand occurs) and inter-demand intervals. The initial demand size and inter-demand interval estimate is calculated as the average quantity over the within-sample (60) observations. The smoothing parameters are not optimised but rather they are fixed to a commonly used value equal to 0.05 (Syntetos and Boylan, 2005). Subsequently, we produce an one-step-ahead estimate for period 61 as well as the forecast error for this period. Then we include the 61st observation in the in-sample block, and we repeat the same process by forecasting for point 62. This process is

repeated until all out-of sample points are exhausted, resulting in 24 one-step-ahead errors.

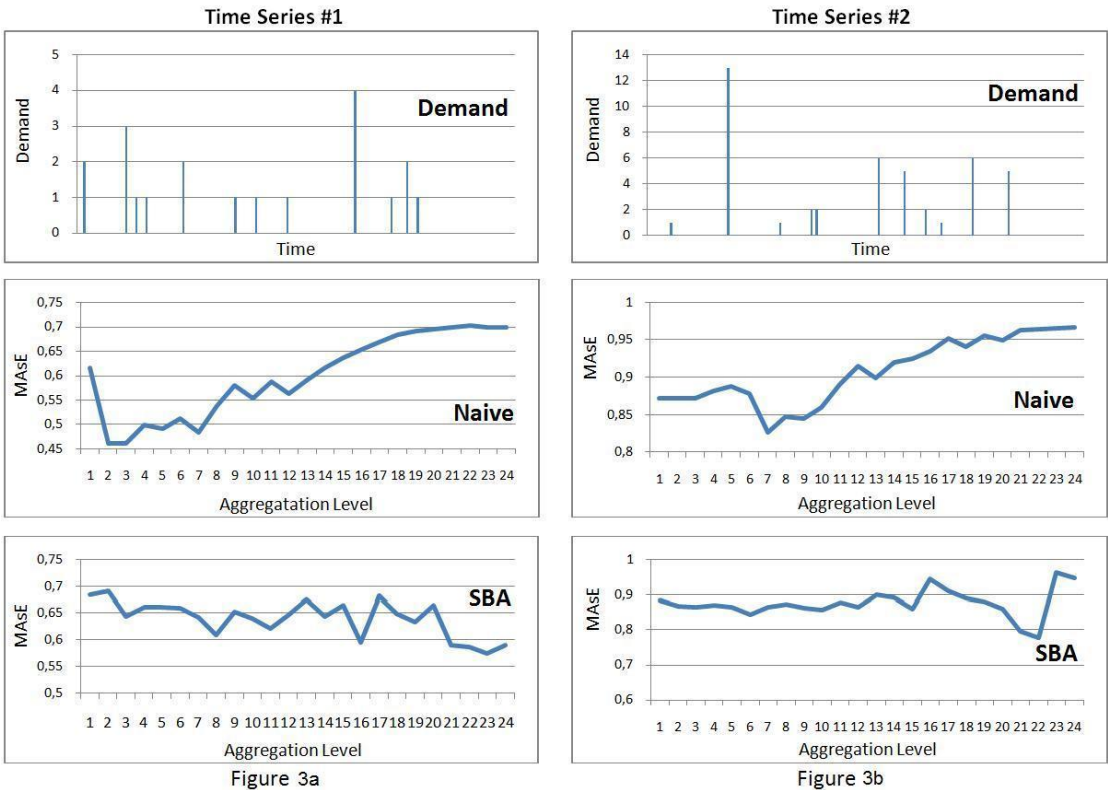
Aggregation level = 2: Starting from the 60th monthly observation, we sum observations backwards in buckets of two (2), resulting in a bi-monthly series consisting of 30 aggregate observations. Subsequently, we create an one-step- ahead bi-monthly forecast (covering periods 61 and 62) using the Naïve method for each series. We disaggregate this bi-monthly forecast into two equal monthly forecasts that correspond to months 61 and 62; the latter forecast is not further utilised since there is only one actual observation held out (for the 61st month). This is the ADIDA(2, Naïve, EQW)¹ forecasting process, where ‘EQW’ denotes ‘equally weighted’. We repeat exactly the same process using the SBA instead of the Naïve method. The process is termed: ADIDA(2, SBA, EQW). Subsequently, we summarise the out-of-sample one-step-ahead errors as discussed above.

Aggregation level = 3...24. Similarly, the experiment continues with time buckets up to 24 periods. At this level there are only two aggregate bi-yearly observations (2 x 24 = 48 observations), so 12 monthly observations at the start of the series remain unused.

The results presented in Figure 3 indicate that the ADIDA process functions as a self-improving mechanism for both estimators. Across all series, the benefit is perhaps more marked for the Syntetos-Boylan Approximation; please also note that for presentation purposes the *y-axis* scales used for each series in Figure 5 are different for the SBA and Naïve estimator. One might have expected more modest (comparative) improvements for

¹ For presentation purposes, and for the remainder of the paper, the ADIDA process will be denoted by: **ADIDA** (*aggregation level, extrapolation method, disaggregation method*).

the former method, since it has in fact been constructed for application on intermittent series. However, the sparseness of data, i.e. the great number of zero observations present in each series (please refer also to the inter-demand interval descriptive statistics in Table 1) renders the Naïve method a very accurate estimation procedure.



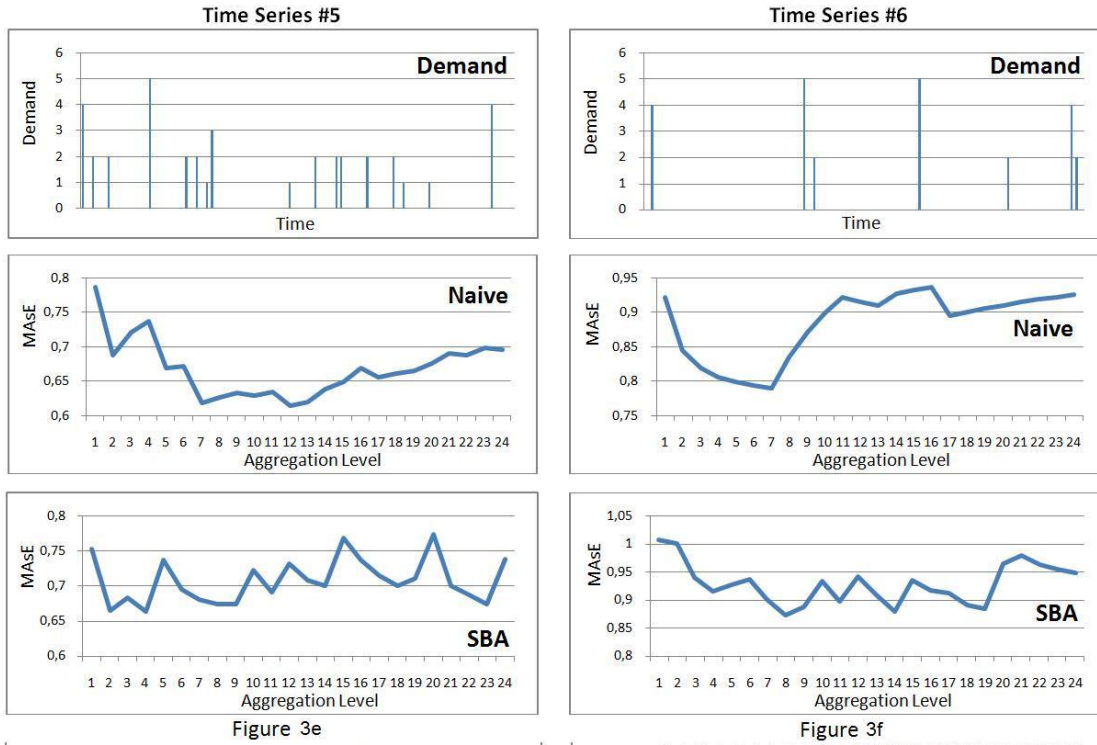
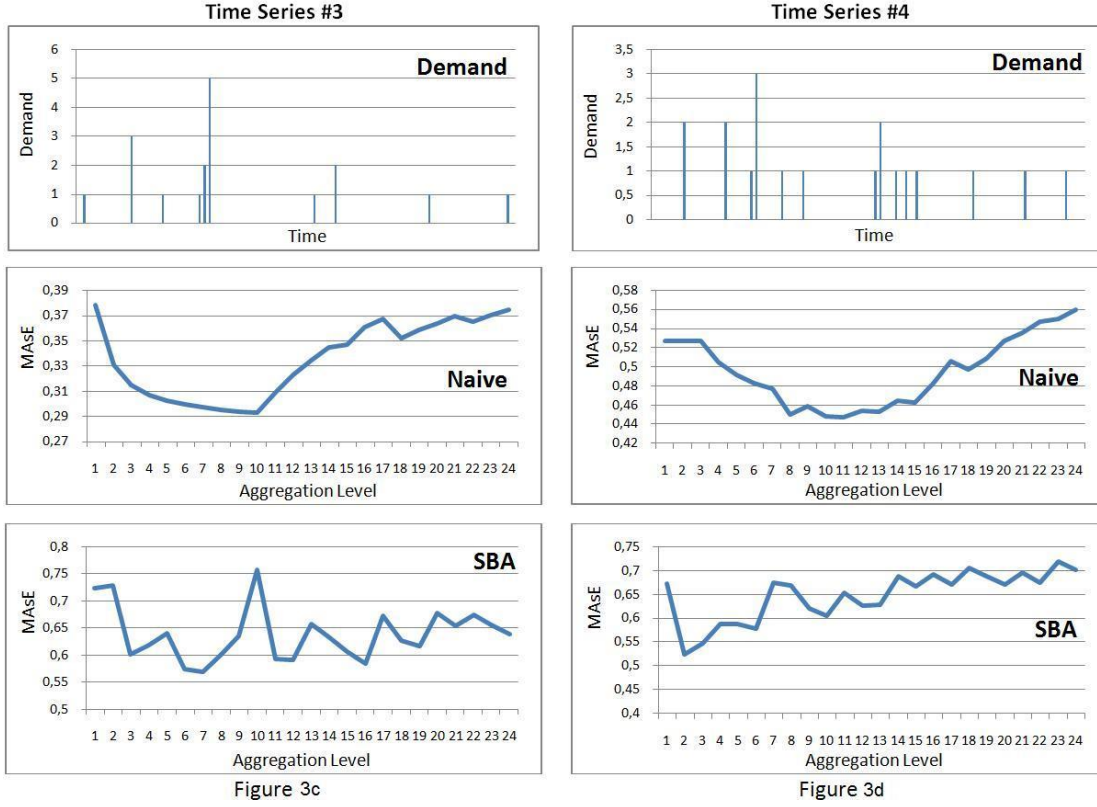


Figure 3. Accuracy results for six randomly selected series

The results also demonstrate that although there is an ‘optimum’ level of aggregation this is not the same across series. This was expected since, theoretically, such a level relates to the underlying demand structure of the series. This issue is further discussed in the last section of the paper where the natural next steps of research are identified. In the following sub-section we investigate the behaviour of the ADIDA framework across the entire dataset considered for the purposes of our experiment.

5.2. Optimal aggregation levels (across series): an empirical determination

Should the dataset be ‘homogeneous’ in nature, an empirical ‘optimum’ aggregation level would be expected across the entire dataset. In the following figure we report the performance of the ADIDA framework on the entire dataset (5,000 SKUs). The experimental structure is identical to that discussed in the previous sub-section, resulting in an evaluation over 120,000 errors (per method, per aggregation level): 24 errors x 5,000 series.

In Figure 4 we see an interesting result. For the Naïve method, a minimum error is achieved via an aggregation level of nine periods. Of course this empirical minimum relates to the specific dataset used for experimentation purposes and the finding may not be necessarily generalised to other situations. In addition, and as discussed in the previous sub-section, the ‘optimal’ aggregation level is different for each individual series. However, this is a promising result in terms of potentially introducing operationalised rules for an entire group of SKUs as well as linking temporal aggregation to cross-sectional issues. This is further discussed in the last section of the paper. The results are similar for the SBA estimator.

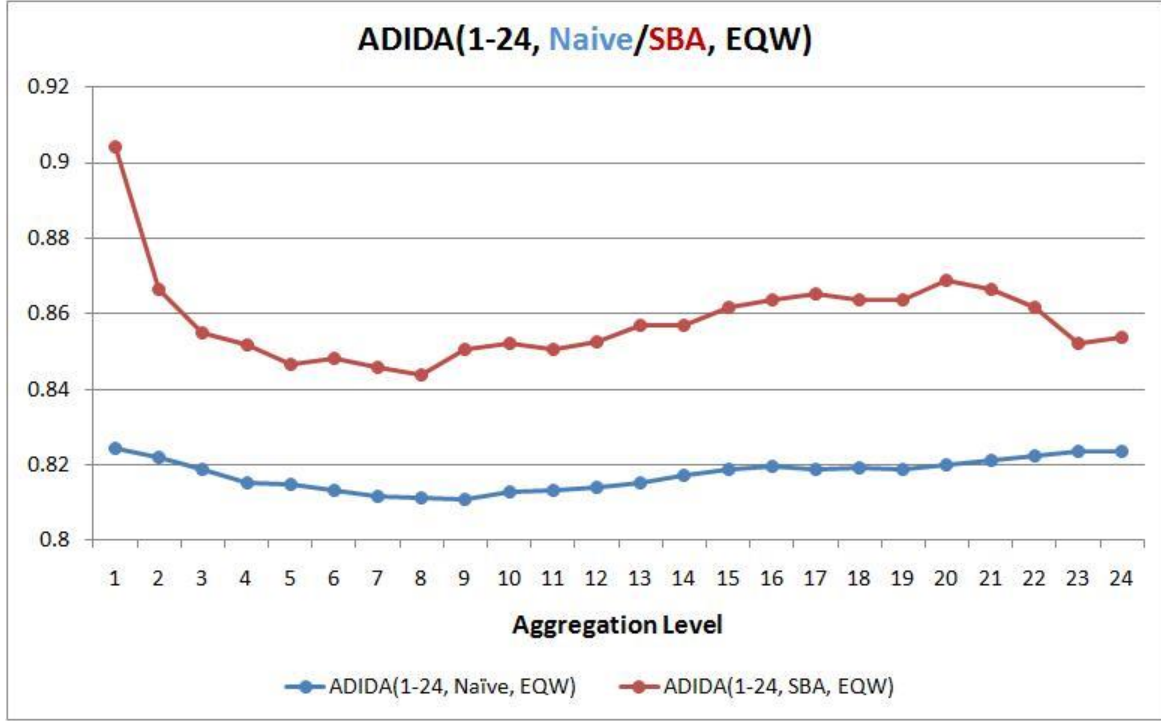


Figure 4. Effect of aggregation level on forecast accuracy for the Naïve and SBA method (across 5,000 series)

The results presented in the previous paragraphs indicate that the ADIDA process may lead to substantial improvements in a single method's application. (In the case of SBA the improvements are also statistically significant at the 5% level².) That is to say ADIDA may be perceived as a *method self-improvement process*, both on individual series and on the entire dataset. The validity of the above results has been further examined and confirmed through the application of two (2) more error measures: Mean Square Error (MSE; which relates directly to variance) and Relative Geometric Root Mean Squared Error (RGRMSE; Syntetos and Boylan (2005) - that has been shown to be very robust on intermittent data).

² 95% confidence intervals were constructed at aggregation level = 1 through the calculation of the sample standard errors (standard deviation of errors / square root of number of errors considered) for both methods. Subsequently, all average errors at the various aggregation levels (> 1) indicated in Figures 6 and 7 (for the Naïve and SBA estimator respectively) were evaluated as to whether they constitute statistically significant improvements/reductions.

5.3. Empirical determination of the best disaggregation method

The simplest possible disaggregation mechanism (and the one employed for the purposes of our research) is that related to using Equal Weights (EQW). That is, for example, if a quarterly forecast is meant to be broken down to monthly ones, each monthly forecast will be equal to $1/3$ of the quarterly one. Furthermore, two other weighing methods were tested without much success, as illustrated in figures 5 and 6. PRW (Previous Weights) is using fractions for the point forecasts that resemble the empirical ones observed in the very last bucket of the original time-series; AVW (AVerage Weights) is using the averages of all the empirical fractions observed throughout all the available historical time buckets of the original time-series.

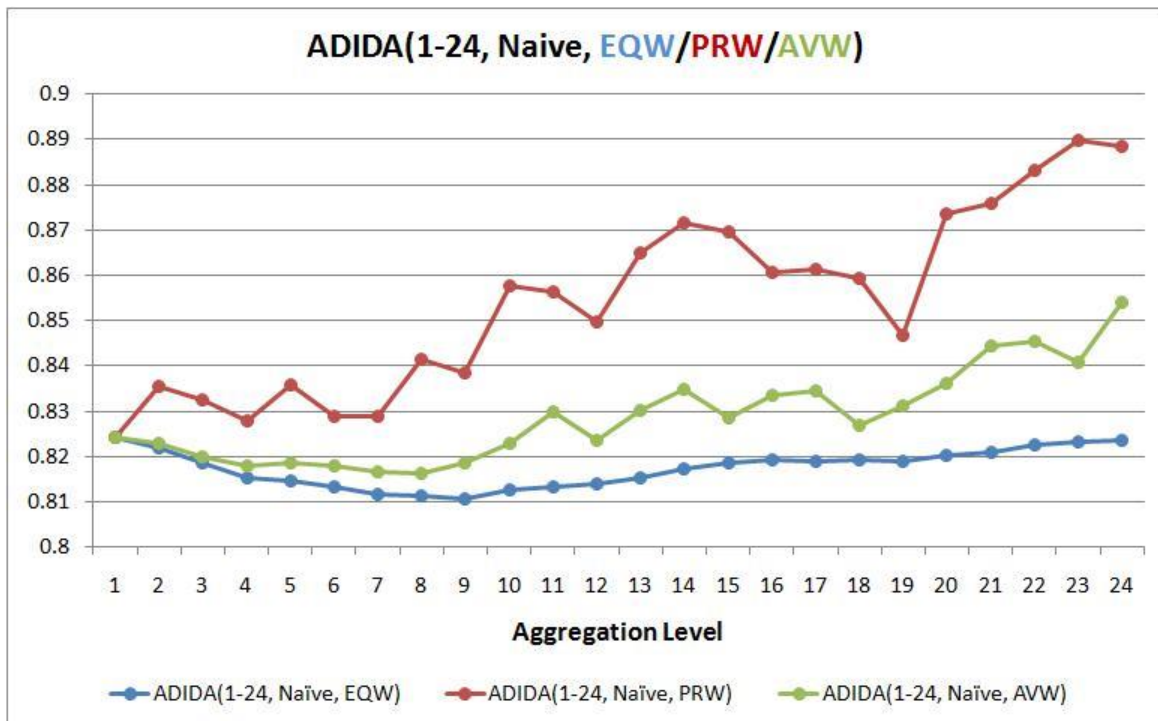


Figure 5. Optimum aggregation level for various disaggregation mechanisms - Naive

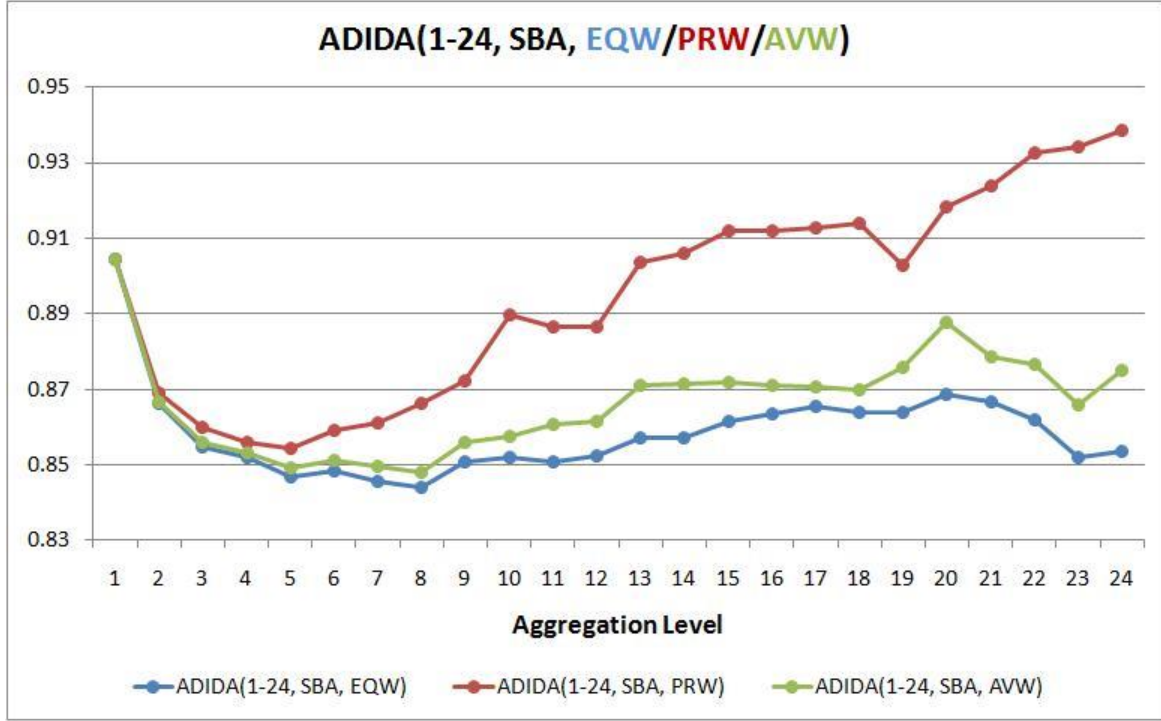


Figure 6. Optimum aggregation level for various disaggregation mechanisms - SBA

The results presented in Figures 5 and 6 are produced through the application of the MASe error measure. The results indicate the ‘best’ performance of the EQW disaggregation mechanism. Similarly to the analysis conducted in the previous sub-section the experiment discussed above was replicated by considering two more accuracy measures: MSE and RGRMSE. The results confirm, overall, the superior performance of EQW. This superiority was theoretically expected due to the stationary nature of the demand data examined in this research (please refer to Section 4). If the data exhibited strong auto-correlation then the PRW and AVW approaches would be expected to perform better.

5.4. Aggregation level = $LT+R$: a managerial-driven heuristic

From the previous sub-section it has become apparent that particular forecasting methods perform better via the ADIDA process and that optimal aggregation levels may be found. The theoretical determination of such a level is left for future research. In this sub-section we are concerned with the specification of the aggregation level based on operational considerations rather than the satisfaction of certain ‘optimality’ conditions.

In particular, it is true to say that in a practical inventory setting, it would make sense to set the aggregation level equal to the lead time length plus one review period, since cumulative forecasts over that time horizon are required for stock control decision making. Such a process renders disaggregation redundant. In Table 2 we present empirical results for an aggregation level = $LT + 1$ (the review period has been set equal to a single period for all SKUs).

Not all 5,000 series were considered in this part of the analysis, as we excluded all series characterized by lead-time (LT) equal to zero (i.e. $LT + R = 1$). Aggregation may well be performed in these series; however here our analysis focuses on the value of aggregation for time buckets that equal $LT + R$, in which case, and for the series under concern, aggregation introduces no different series than the original ones. In addition, all series characterized by: $LT + R > 24$ were also excluded in order to enable a sufficient out-of-sample sub-set to be produced (please see experimental details below). This elimination process resulted in 4,352 SKUs considered for simulation purposes.

We considered the Naïve method and SBA and used as the holdout period the last two $LT + 1$ cumulative periods of each series. We performed a sliding simulation over this out-of-sample via producing two $LT + 1$ steps-ahead forecasts; thus we calculated two $LT + 1$ steps-ahead errors (*CumError*) for these periods:

$$CumError = \text{Cumulative Demand over } LT + 1 \text{ periods} - \text{Forecast over } LT + 1 \text{ periods}$$

as series have been aggregated in buckets of $LT + 1$. Absolute scaled errors were subsequently calculated; for scaling purposes we used the in-sample average $LT + 1$ steps ahead forecasts. The mean and Median values of this metric (*MAsE* and *MdAsE* respectively) are reported in Table 2. Three more error measures were considered: Bias (Mean and Median Signed Error: denoted by *ME* & *MdE* respectively, error = actual - forecast), MSE and RGRMSE.

| 4,352 SKUs | | Forecasts | | ADIDA Forecasts | |
|-----------------|---------------|-----------|---------|-----------------|---------|
| | | Naïve | SBA | Naïve | SBA |
| Bias | <i>ME</i> | 2.35 | -3.57 | -0.39 | -2.55 |
| | <i>MdE</i> | 1.00 | -1.59 | 0.00 | -1.37 |
| Scaled Errors | <i>MAsE</i> | 125.53% | 92.13% | 99.84% | 89.24% |
| | <i>MdAsE</i> | 13.04% | 20.93% | 19.56% | 19.65% |
| Squared Errors | <i>MSE</i> | 8147.29 | 2082.39 | 3092.99 | 2084.51 |
| Relative Errors | <i>RGRMSE</i> | 1.00 | 0.55 | 0.83 | 0.52 |

Table 2. ADIDA forecasts evaluation - Cumulative forecast horizon = $LT + 1$

- **Bias:** ADIDA($LT+1$, Naïve, EQW) has the lowest bias with the Naïve method being the second best. SBA performs slightly worse, although its performance is improved via ADIDA. The negative signs of the Median Errors of SBA indicate

that there are some positive outliers, where forecasts seriously under-estimate the actual.

- **Scaled Errors:** SBA as expected is the most accurate forecasting approach. Furthermore, ADIDA(LT+1, SBA, EQW) reduces the *Mean Absolute Scaled Error* by an absolute value of 2.89%, which could be seen as a percentage improvement of 3.13%. Median values for the metric are much lower indicating the existence of outliers.
- **Squared Errors:** indicative of the volatility of the provided forecasts. SBA has the lower value and it practically remains unchanged via the ADIDA process, while the impact in the case of Naïve forecasts is substantial (reduced to the 37% of the original *MSE*).
- **Relative Errors:** this relative (to Naïve) metric (*Relative Geometric Root Mean Squared Error*) confirms the results obtained by the aforementioned error measures.

6. Conclusions and extensions

Aggregation is an appealing strategy for intermittent demand because such items are often voluminous and have good lengths of demand history. Moreover, most inventory forecasting software packages support aggregation of data. Although this would typically cover cross-sectional aggregation (i.e. aggregation across series) minor customization should render temporal aggregation and ADIDA a feasible strategy for many organizations. Until now, most packages have not fully exploited the richness of data that is available, relying instead on extrapolation of a single item, with no temporal aggregation.

In this paper we have been concerned with empirically analysing the effects of temporal aggregation for forecasting intermittent demand requirements. We did so by means of experimentation on 5,000 SKUs from the RAF. The empirical results indicate that our proposed methodology may indeed offer considerable improvements in terms of forecast accuracy. Forecast variance reductions have also been demonstrated through the consideration of the MSE metric. The main findings of this study are the following:

- The ADIDA process may lead to substantial improvements in a single method's application; thus, it may be perceived as a *method self-improvement mechanism*.
- The empirical results demonstrate that an optimal aggregation level may exist. This is true both at the individual series level and across series (i.e. for the entire dataset).
- Setting the aggregation level equal to the lead time length plus one review period, shows very promising results. This simple heuristic would make sense in a practical inventory setting, where cumulative forecasts over that time horizon are required for stock control decision making.

Originally, our motivation for experimenting with the proposed methodology related to the introduction of methods designed for fast moving items into the task of forecasting intermittent demand requirements. However, and as discussed above, the results indicate that *aggregation* constitutes a self-improvement mechanism for intermittent demand estimators, such as the SBA. Further research should involve the actual application of methods so far tested only in regular demand series (such as Damped Exponential

Smoothing and the Theta model, Assimakopoulos and Nikolopoulos, 2000), on aggregate series resulting from intermittent data. In addition, replication of this study on other datasets should also prove to be a valuable exercise.

We have recently argued (Syntetos *et al*, 2009b) that in an inventory forecasting setting extrapolation methods should not only be evaluated with respect to their forecast accuracy but also in terms of their stock control implications, as measured through accuracy implication metrics (such as inventory costs and service levels achieved). Exploring the effects of temporal aggregation on stock control is an interesting line of further research and certainly worthwhile pursuing from a practitioner's perspective. In addition, and given the relevance of aggregating demand in time buckets that equal the lead time plus review period to stock control, more research in this area would appear to be merited.

The interaction between temporal and cross-sectional forecasting is also an exciting area of research. The latter may be 'product'/SKU or location-related and empirical work could be undertaken to examine the potential benefit of combining these various forms of aggregation, both in terms of forecast accuracy and inventory control performance.

Finally, and most importantly, in this paper we have not considered the theoretical underpinnings of the ADIDA process. Theoretical MSE expressions, along the lines discussed by Syntetos *et al* (2005) should enable: i) the identification of optimum levels of aggregation and their linkage to the series' underlying properties; ii) contrasting the performance of various estimators when applied on aggregated data for the purpose of choosing one, and iii) the development of demand classification theory for forecasting

(and stock control) purposes. Previous work in this area has modelled demand as a Bernoulli process and relied upon three key parameters: i) average inter-demand interval (or correspondingly the Bernoulli probability of demand occurrence); ii) the mean demand size (when demand occurs), and iii) the variance of demand sizes. One opportunity for extending the work discussed above is through the consideration of the *aggregation bucket length* and the *length of the series* (as two additional parameters). Alternatively, Poisson-based modelling constitutes also a natural extension.

References

- Assimakopoulos V and Nikolopoulos K (2000). The Theta model: a decomposition approach to forecasting. *International Journal of Forecasting* **16**: 521-530.
- Boylan JE, Babai MZ, Syntetos AA and Smith L (2008). Bootstrapping for intermittent demands: theory and practice. The 15th *International Symposium on Inventories*. Budapest, Hungary.
- Brännäs K, Hellström J and Nordström J (2002). A new approach to modelling and forecasting monthly guest nights in hotels. *International Journal of Forecasting* **18**: 19-30.
- Croston JD (1972). Forecasting and stock control for intermittent demands. *Operational Research Quarterly* **23**: 289-303
- Hyndman RJ and Koehler AB (2006). Another look at measures of forecast accuracy. *International Journal of Forecasting* **22**: 679-688.
- Johnston FR and Harrison PJ (1986). The variance of lead-time demand. *Journal of the Operational Research Society* **37**: 303-308.
- Johnston FR, Boylan JE and Shale EA (2003). An examination of the size of orders from customers, their characterization and the implications for inventory control of slow moving items. *Journal of the Operational Research Society* **54**: 833-837.
- Porras EM and Dekker R (2008). An inventory control system for spare parts at a refinery: An empirical comparison of different reorder point methods. *European Journal of Operational Research* **184**: 101-132.
- Silvestrini A and Veredas D (2008). *Temporal aggregation of univariate and multivariate time series models: A survey*. Temi di discussione (Working papers) 685 (August 2008), Banca

d'Italia.

- Snyder RD, Koehler AB and Ord K (1999). Lead time demand for Simple Exponential Smoothing: an adjustment factor for the standard deviation. *Journal of the Operational Research Society* **50**: 1079-1082.
- Syntetos AA, Babai MZ, Dallery Y and Teunter R (2009a). Periodic control of intermittent demand items: theory and empirical analysis. *Journal of the Operational Research Society* **60**: 611-618.
- Syntetos AA and Boylan JE (2001). On the bias of intermittent demand estimates. *International Journal of Production Economics* **71**: 457-466.
- Syntetos AA and Boylan JE (2005). The accuracy of intermittent demand estimates. *International Journal of Forecasting* **21**: 303-314.
- Syntetos AA and Boylan JE (2006). On the stock-control performance of intermittent demand estimators. *International Journal of Production Economics* **103**: 36-47.
- Syntetos AA, Boylan JE and Croston JD (2005). On the categorization of demand patterns. *Journal of the Operational Research Society* **56**: 495-503.
- Syntetos AA, Boylan JE and Croston JD (2006). A note on the categorization of demand patterns. *Journal of the Operational Research Society* **57**: 1257-1258.
- Syntetos AA, Nikolopoulos K and Boylan JE (2009b). Judging the judges through accuracy-implication metrics: the case of inventory forecasting. *International Journal of Forecasting*, in press.
- Teunter R and Sani B (2009). On the bias of Croston's forecasting method. *European Journal of Operational Research* **194**: 177-183.
- Tsai H and Chan KS (2005). Temporal aggregation of stationary and non-stationary discrete-time processes. *Journal of Time Series Analysis* **26**: 613-624.
- Willemain TR, Smart CN, Shockor JH and DeSautels PA (1994). Forecasting intermittent demand in manufacturing: a comparative evaluation of Croston's method. *International Journal of Forecasting* **10**: 529-538.
- Willemain TR, Smart CN and Schwarz HF (2004). A new approach to forecasting intermittent demand for service parts inventories. *International Journal of Forecasting* **20**: 375-387.

